CHARACTERIZATION OF *BSA*, A BOVINE CLASS I MHC SEQUENCE

An Undergraduate Research Scholars Thesis

By

VIVEK AJAY KOSHTI

Submitted to Honors and Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor: Dr. Loren Skow

May 2014

Major: Biomedical Sciences

TABLE OF CONTENTS

Page

ABSTRACT

Characterization of *BSA*, a Bovine Class I MHC Sequence. (May 2014)

Vivek Ajay Koshti
Department of Biomedical Sciences
Texas A&M University

Research Advisor: Dr. Loren Skow
Department of Veterinary Integrative Biosciences

The class I MHC region in the vertebrate genome contains genes for receptor proteins that present cytosolic antigens to lymphocytes in the initial phase of the immune response. Most cattle express a class I gene called "*BSA*" as the most prevalent mRNA transcript in leukocytes. However, sequence corresponding to *BSA* is missing from the bovine reference genome sequence and from genomic libraries prepared from several different cattle. We used locus specific primers to amplify the *BSA* gene in the DNA from L1 Domino, the donor for the CHORI240 BAC library and from each of the donors for two other genomic libraries. Additionally, we characterize the sequence variation of *BSA* among 92 representative BoLA haplotypes, including an evolutionary-conserved, trans-species frameshift polymorphism in cattle and bison. Variation at the *BSA* locus should be included in the design of SNP arrays and alignments to reference genomes for the identification of BoLA-associated phenotypes in genome wide association studies (GWAS).

DEDICATION

This thesis is dedicated to my parents, Ajay and Anjali Koshti, whose support and guidance has allowed me to develop as a person and pursue my goals.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Loren Skow and Erika Downey for their direction, patience, and enthusiasm in my investigation.

NOMENCLATURE

BoLA	Bovine Leukocyte Antigen
MHC	Major Histocompatibility Complex
GWAS	Genome-wide association study

CHAPTER I

INTRODUCTION

The major histocompatibility complex (MHC) is a critical region of mammalian genomes for establishing initial host responses to infection. The MHC contains numerous genes crucial for initiating the innate and adaptive immune responses in animals exposed to pathogens and vaccines. Not surprisingly, the MHC has been associated with more diseases and disease susceptibilities in humans and model organisms than any other region of the genome (1). Similarly, polymorphisms in the bovine MHC (BoLA) are associated with susceptibility to many infectious diseases and parasites, including mastitis, dermatophilosis, tick resistance, bovine leukemia virus (BLV) induced leukemia (2-7) and vaccination response (8, 9). Genes in the BoLA class I region function to present viral and "self-antigens" to the cell mediated branch of immune response. These genes evolve by a "birth and death" process (9) such that individual genes may be duplicated or deleted by unequal recombination. Consequently, different haplotypes of BoLA demonstrate copy number variation (CNV) and may contain different numbers of class I genes.

Genes of the MHC class I are among the most highly polymorphic genes in mammalian genomes, a characteristic that can confound genetic analysis. We utilize animals identical (homozygous) at BoLA to characterize variation in *BSA* without confounding variation present in heterozygotes. Dr. Skow's laboratory has accumulated samples of DNA from homozygote cattle representing 92 different haplotypes of BoLA. This DNA archive represents the largest known collection of BoLA haplotypes and provided the principal experimental material for this study.

Expression studies (10) identified a class I mRNA designated "*BSA*" present at high levels in white blood cells of an Angus bull (10) However, we discovered that the *BSA* sequence is missing from the assembly and trace files of the Bovine Reference genome sequence, from the bovine BAC library CHORI 240, and from phage and BAC genome libraries produced at Texas A&M University. Consequently, the gene markers for *BSA* and other nearby genes are absent from SNP arrays and reference assemblies used to search for associations between genetic variation and health phenotypes in cattle. In this study, we demonstrate the presence of *BSA* sequence in DNA from L1 Domino, donor for the CHORI 240 BAC library, and the donors of the TAMU genome libraries. We also describe *BSA* sequence variation among the BoLA haplotypes that may provide a better understanding of genomics for improved cattle health and food safety.

CHAPTER II

METHODS

All experiments were conducted in the laboratory of Dr. Loren Skow, Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences. Dr. Skow's laboratory is located in Room 335 of the Veterinary Medical Research Building.

Locus–specific PCR primers, Bucon New 5'-CCGCCCAACTCTGCTTCTTTC-3' and Blcon 5'-TCCTTCCCGTTCTCCAGGTATC-3',were used to determine the presence or absence of *BSA* in DNA samples previously extracted from each of the 92 BoLA homozygous cattle. These primers, which hybridize to conserved locus-specific sequences specific for *BSA*, were designed in the Skow lab to specifically amplify *BSA* sequences using the polymerase chain reaction (PCR). Each PCR reaction contained 50-100 ng of DNA, 1x PrimerSTAR buffer (TaKaRa PrimeSTAR), 0.3 pmol each primer, 0.25 mM dNTPs (each), and 1.25 units PrimeSTAR GXL polymerase (TaKaRa). TaKaRa high fidelity thermostable polymerase was used to minimize experimental artifacts in the amplification products which were cloned and sequenced. Amplification parameters include 30 cycles of 98°C for 10 sec, 60°C for 15 sec, and 68° for 80 sec. Presence or absence of amplicons was determined by electrophoresis on 1.0% agarose gels containing 0.005% ethidium bromide and photographed under UV illumination. Positive and negative controls were included in each amplification experiment. Amplicons of *BSA* were extracted from the gel and purified using a Qiagen Gel Extraction Kit (Qiagen®). Purified DNA was quantified using electrophoresis on a 1.0% agarose gel containing 0.005% ethidium bromide. DNA was frozen at -20°C until cloning.

The PCR reaction was used to restore 3' A (adenosine) overhangs to the gel purified DNA to facilitate cloning. The 13ul PCR reaction contained 10-100ng of gel purified DNA, 1x buffer (Sigma Aldrich), 8mM dNTPs, and 1 unit JumpStart REDTaq polymerase (Sigma Aldrich), incubated at 72°C for 10 min. Purified PCR products were cloned into the pCR™4-TOPO® Vector (Invitrogen) according to manufacturer's recommendations. Three microliters of pCR 4-TOPO vector were added to 50 ul suspension of chemically competent *E. coli* cells (Invitrogen), incubated on ice for 30 min, and heat shocked at 42°C for 30 sec. 200 ul SOC media (Invitrogen) was added to each tube and incubated at 37°C with shaking for one hour. One hundred microliters of transformed *E. coli* cells from each PCR sample were plated on petri dishes containing Luria Broth agar with 75ug/ml ampicillin to select for plasmid containing colonies. After incubation at 37°C overnight, we picked five clonally isolated colonies separately into 10 ml of ampicillin/LB broth and grew the cultures to an OD620 ~0.8. Glycerol stocks (12.5%) were made from each culture and 100 ul of each clone was removed and transferred to a 96 well plate for shipment to Beckman Coulter for sequencing. The remaining glycerol stocks were stored frozen at -80 C.

Beckman's Coulter Genomics (Danvers, MA) purified plasmid DNAs from the 12.5% glycerol preparations and conducted bidirectional dideoxy sequencing using M13 Forward 5'-GTTTTCCCAGTCACGAC-3' and M13 Reverse 5'-CAGGAAACAGCTATGAC-3' as sequencing primers. Sequencing data were manually trimmed, edited, and consensus sequences assembled using Sequencher and CLC Genomics Workbench 6. Consensus sequences were aligned against the NCBI *Bos bovis* nucleotide database using the BLASTN program.

For phylogenetic analysis, we trimmed consensus *BSA* sequences to a common length and aligned them using CLC Genomics Workbench 6. Relationships among the *BSA* alleles were inferred by constructing a phylogenetic tree using UMPGA with 10,000 bootstrap iterations. Three different MHC class I bovine sequences (*BSN*, *BSX*, and *BSC*) and the human *HLA-A* class I sequences were used as outgroups to root the tree.

CHAPTER III

RESULTS

BSA was successfully amplified in 60.8% (76/125) of the BoLA homozygous animals. Sequences of 43 of the 62 unique alleles identified among the 125 homozygotes represent newly described alleles; 19 of the sequences aligned perfectly to sequences previously deposited in NCBI. Additionally, we demonstrated that *BSA* could be amplified from DNA of Domino (donor for the CHORI 240 BAC library) and from the donor animals of the TAMU genomic libraries. All but one of the 62 unique sequences aligned with >99% identity to AF396750, the original published sequence of *BSA*. The single exception is a sequences from an animal of the Romosinuano breed, which aligned with ~94% identity to AF396750.1.

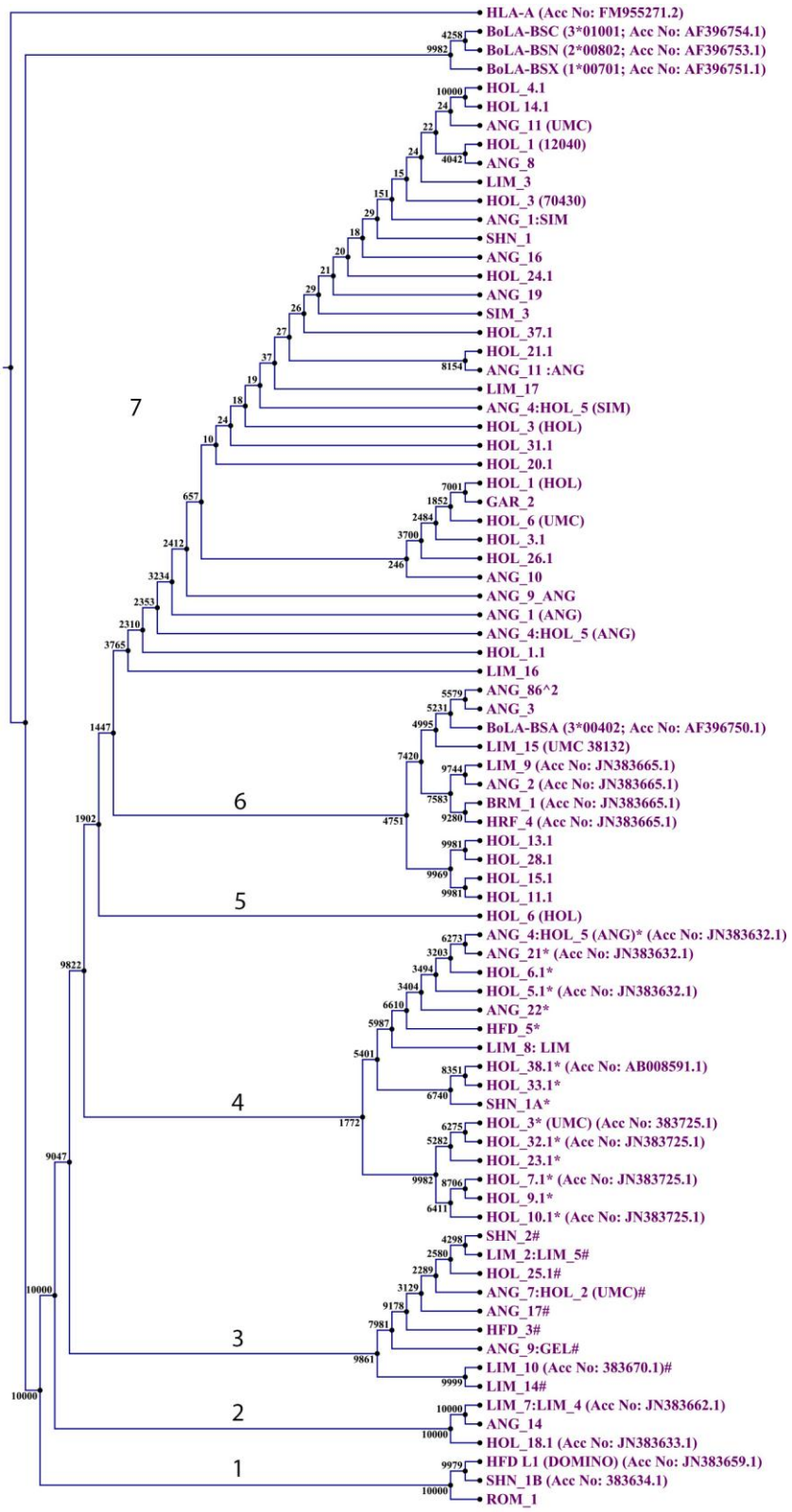

Figure 1. Phylogenetic analysis of *BSA* alleles. UPGMA rooted tree of 62 *BSA* alleles (Exons 1 –3) in 66 different SNP-defined BoLA homozygotes. * denotes those samples with the 2 bp deletion in exon 2. # denotes samples that had a 21 bp deletion in intron A.

Based on the phylogenetic analysis (Figure 1), the 62 *BSA* alleles clustered into at least seven major groups. Bootstrap values strongly supported the relationships of groups 1 - 6, whereas support was weak for the sequences in group 7. Sequences in group 4 contain a two base pair deletion (Figure 1), which causes a frameshift mutation in exon two and premature stop codon in exon three. Consequently, these genes encode a soluble class I protein (11). Fifteen of the *BSA* sequences (24.2%) contain this deletion, and 10 out of the 15 sequences are from Holstein samples. A second deletion of 21 base pairs in intron A was identified in eight *BSA* alleles (group ; Figure 1).

CHAPTER IV

CONCLUSION

The DNA from Domino, the donor for the CHORI 240 BAC library, contains a BoLA-*BSA* sequence as do the donors of two TAMU genomic libraries. For unknown reasons, none of the three genomic libraries contain the corresponding *BSA* sequence. *BSA* could also be amplified in approximately 60% of the 92 different BoLA haplotypes. This result is not surprising, as no single class I locus has been shown to be present in all defined BoLA haplotypes (12). Variation at the *BSA* locus may be important for identification of BoLA-associated health phenotypes in genome-wide association studies and should be included in design of new microarrays and additional genome sequence assemblies for cattle to provide a complete assay for future genome-wide association studies.

REFERENCES

1. Thorsby, E. Invited anniversary review: HLA associated diseases. Hum Immun 53:1, 1997.

2. Park YH, Joo, YS, Park, JY, Moon, JS, Kim, SH, Kwon, NH, Ahn, JS, Davis, WC, and Davies, CJ. Characterization of the lymphocyte subpopulations and major histocompatibility complex haplotypes of mastitis-resistant and susceptible cows. J. Vet. Science 5(1):29-39, 2004.

3. Sharif, S, Mallard, BA, Wilkie, BN, Sargeant, JM, Scott, HM, Dekkers, and Leslie JC. Associations of the bovine major histocompatibility complex DRB3 (BoLA DRB3) with production traits in Canadian dairy cattle. Anim. Genet. 30(2):157-60, 1999.

4. Maillard JC, Berthier D, Chantal I, Thevenon S, Sidibé I, Stachurski F, Belemsaga D, Razafindraïbé H, and Elsen JM. Selection assisted by a BoLA-DR/DQ haplotype against susceptibility to bovine dermatophilosis. Genet Sel Evol 35:S193-200, 2003.

5. Ballingall KT, Luyai A, Rowlamnds GJ, Sales J, Misoke AJ, Morzaria SP, and McKeever DJ. Bovine leucocyte antigens major histocompatibility complex class II DRB3*2703 and DRB3* 1501. Infect Immun. 72(5):2738-41, 2004.

6. Mirsky ML, Olmstead C, Da Y, and Lewin HA. Reduced bovine leukaemia virus proviral load in genetically resistant cattle. Anim. Genet. 29(4):245-52, 1998.

7. García-Briones MM, Russell GC, Oliver RA, Tami C, Taboga O, Carrillo E, Palma E, Sobrino F, and Glass EJ. Association of bovine DRB3 alleles with immune response to FMDV peptides and protection against viral challenge. Vaccine 8;19(9-10):1167-71, 2000.

8. Glass EJ. Genetic variation and responses to vaccines. Anim Health Res Rev 5(2):197-208, 2007.

9. Ellis SA and Ballingall KT. Cattle MHC: evolution in action? Immunol Rev Feb;167:159-68, 1999.

10. Garber T, Hughes A, Watkins D, and Templeton J. Evidence for at least three transcribed BoLA class I loci. Immunogen 39:257-265, 2004.

11. Ramlachan N, Womack JE, and Skow LC. Analyses of bovine major histocompatibility (BoLA) class I loci reveal a novel frameshift mutation in a BoLA-A allele prevalent in wild and feral bovids. In: 12th International Congress of Immunology and 4th Annual Conference of FOCIS, Montreal, Canada, 2004.

12. Ellis SA, Holmes EC, Staines KA, Smith KB, Stear MJ, McKeever DJ, MacHugh ND, and Morrison WI. Variation in the number of expressed MHC genes in different cattle class I haplotypes. Immunogen 50:319-328, 1999.